REX

JUSTIN

1. Clustering
   1. We see structure

Asymmetry

* 1. What is the cluster at the center of the arms/centrally connected population?
  2. Functional circuits?

1. Hubs
   1. Operational definition for hubs- how did we define them?
      1. Show degree distribution w/ dotted line representing mean jumps in degree (difference matrix) + 1SD
   2. Identify, talk about function, birth time, position in worm body
   3. Show different analyses
      1. Hub similarity matrices
         1. Similar ones seem to be actual pairs
2. Topological importance of hubs: hubs have been found previously but importance for network has not been demonstrated
   1. Show power law distribution of hubs, suggesting scale-free behavior.
      1. But after just 3 top hubs deleted in serial order, the distribution is ruined and everything is equally likely to be classified as a hub due to even distribution of degrees
   2. How important are hub neurons for network structure? Re-clustering when they are deleted.
   3. What happens to functional circuits identified in (1) when they are deleted?
   4. Figures
      1. Clustering plots before and after
      2. Hub plots (connection diagrams) before and after

Spectral Clustering

## Motivation

Modelling of neuron networks as graph allows us to analyze various neuron circuits that corresponds to different functional circuits of the organism. In recent researches people have discovered the neuron circuits that corresponds to various behaviors of C. elegans, and identified functionalities of most of the neurons in C. elegans, though few still remain unknown. Our aim is to establish a general framework that could analyze the structure of neuron network, reproduce and even identify new subgraphs and important nodes within the neuron network with significant functional significance.

A graph that models the neuron network consists of the set of vertices , which are neurons, and edges which are the synapses. The spatial structure of C-elegans (fig) does not reveal the connectivity of neurons. Consequently, to visualize the graph using spatial structure of neurons only reveals few information about significant clusters, hubs and paths. Spectral graph theory comes in to play, because it has both theoretical bases for graph clustering and partitioning, and experimentally out-performs traditional methods, such as Markov Chain Monte Carlo Methods (MCMC)(Sohn et al. 2011) , Expectation Maximization (EM) (Zhang et al. 2002) etc, in terms of efficiency and effectiveness (von Luxburg 2007). Here we attempt to apply this method on the neuron network of C. elegans, from which we show the usefulness of the technique and the insight it gives in understanding the graph model of C. elegans connectome.

## Graph setup

The synapses that form the neuron network of C. elegans are of three types.

* Electrical synapses, also called gap junctions are formed by close contact between cells. In the nervous system, there exist gap junctions between neurons and between muscle cells. The electric synapses are generally considered symmetric: it allows information transfer between neurons in both directions. When reflected in the adjacency matrix of the neuron network graph, the subgraph consisting of all electrical synapses is symmetric.
* Cemical synapses may occur between one presynaptic and one post-synaptic cell (a monad) or more than one post-synaptic partner (a polyad; two recipients make it a dyad and three recipients make it a triad).Chemical synapses are made en passant between neighboring processes where synaptic swellings are formed along the process shafts. (wormatlas) Th subgraph consisting of all chemical synapses is generally asymmetric.
* The neuromascular junction is a special chemical synapse where neuron input to muscles occurs. Since this type of junction does not connect neurons to neurons, we generally exclude them in our connectome analysis unless otherwise stated.

From dataset, we are able to construct the symmetric component of electrical synapses network, and the assymetric component of chemical synapses network. We represented them using sparse matrix form in MATLAB from which first few eigenvectors can be efficiency calculated using iterative methods. The edge weights are defined by the synapse strength available in the dataset (need reference).

For each individual neuron represented as vertices in the graph, we associate them with 3 general types: sensory neurons, which mainly respond to the stimuli of external environment; motor neurons, which have neuronmascular junctions to muscle cells and are in charge of physical activity of the work; and inter-neurons, which mediates between neurons with different types. The type information of vertices is represented by different colors in the visualization, that provides understanding of the general information flow of the neuron network.

## Brief introduction to the technique

Spectral clustering is done by embedding the graph using eigenvectors of graph Laplacian as coordinates. When the eigenvectors corresponding to the smallest non-zero eigenvalues is chosen, the resulting embedding is an approximation to the Ratio Cut problem in graph theory (von Luxburg 2007).

The graph Laplacian is defined as:

Where is the diagonal matrix where , the degree of the ith vertex in the graph.

The Graph Laplacian satisfies the following property, as shown in the appendix, that for any vector of the same dimension as the column vectors of ,

Where is the adjacency matrix of the graph, and is the ith entry of the vector . It is called the quadratic form of a graph. In a special case, when is the normalized eigenvector corresponding to the smallest non-zero eigenvalue of , it follows that

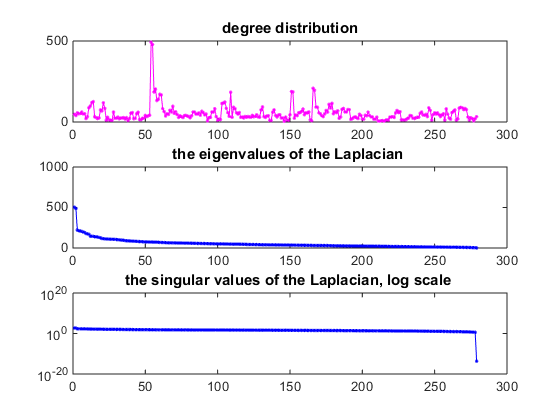
The quadratic form reaches its minimum with the constraint that is perpendicular to the null-space of . Hence the spectral embedding that uses the eigenvectors corresponding to the smallest few non-zero eigenvalues optimizes for the total length of graph connections. If two neurons are closely connected through each other by strong synapses in very few steps, they are embedded close to each other in the resulting embedding.

When the graph is disconnected or almost disconnected (a neuron is connected to rest of the network with very few edges), the multiple eigenvectors correspond to the eigenvalue 0, and these eigenvectors essentially become connectivity vectors. Thus in practice we use the eigenvectors corresponding to the few non-zero eigenvalues of graph Laplacian.

The general procedure is to perform eigen-decomposition on the graph Laplacian generated from adjacency matrix of C-elegans neuron network, and choose eigenvectors corresponding to the smallest few eigenvalues as coordinates for each vertex. Plot the vertices using their respected coordinates. Finally, linear regression and K-means algorithm are performed on the resulting graph embedding, to obtain useful information about the structure of the connectome of C-elegans.

## Results

The degree of each vertex is first calculated, from which we generates the graph Laplacian and perform spectral analysis. As can be seen from the figure, the eigenvalues decrease almost exponentially, except the last eigenvalue which is 0. There is only one eigenvector corresponding to 0 eigenvalue, implying that the neuron network consisting of 279 neurons is connected.



The embedding result shows that there is

## Information flow

In the spectral embedding of unnormalized graph Laplacian, we can see that there are 3 points that are particularly far away from the central cluster of neurons, and 3 other points that are between the previous 3 points and the central cluster.

The 3 neurons corresponding to the faraway 3 points are: AVAR, ABBR, RIAR;

The intermediate 3 neurons are: RIAL, PVCL, DVA.

All of which have common features of being inter-neurons with high connectivity. The hub analysis section addresses the significance of these neurons in more detail.

The normalized version of graph Laplacian gives more interesting results in terms of the structure of connectome. From the embedding, we can see 4 arms that stem from a central cluster that mainly comprises of inter-neurons. The 4 arms have different compositions in terms of neuron types.

In general, the 4 arms represent 4 different set of functionalities of C. elegans. The motor and sensory neurons in each arm seldom communicate directly, but rather will go through an inter-neuron near the center of the embedding. For sensory and motor neurons in charge of similar functionalities, they communicate using a relatively shorter path by going through more peripheral inter-neurons located in the arm that they belong to in the embedding. However, sensory and motor neurons corresponding to different set of functionalities are located in different arms. They communicate with each other by going through a longer path by connecting to a more central command neuron located at the center of the embedding, and then travel to the other arm.

Top arm mainly consists of motor neurons:

Neurons in this arm include DD neurons, which are responsible for s**inusoidal body movement-locomotion.**

DB neurons that are in charge of propagation of rhythmic activities along the body during forward locomotion.

VA is responsible for locomotion.

Almost no inter-neurons are present in this arm. The motor neurons receive their signal primarily through the central cluster of interneurons.

There are some neurons that act both as sensory and motor neurons in this arm. Eg. VB neurons are responsible for both locomotion and proprioception. This compensates for the efficiency of neuron activities without inter-neurons.

Bottom cluster consists mainly of sensory neurons:

Neurons in this arm include PDEL, PDER (basal slowing response); PHEL, PHER are responsible for thermal sensory neurons.

The important interneuron in this arm is DVA, one of the less important hub neurons identified: it provides input to both the anterior and posterior touch circuits. Animals lacking these neurons respond to tap stimulus with diminished forward accelerations and reversals.

Left cluster has similar amount of sensory and motor neurons

The important hub neurons in this arm include RIAL, RIAR. They are "second layer"interneurons in the process of integration of information from the outside world and the inner state of the animal, which then leads to a behavioral response.

Right arm has similar amount of sensory and motor neurons.

The important hub neurons in this arm include PVCL, PVCR. They function as command interneuron specifically for**forward locomotion;** drives forward movement of the animal along with [AVB](http://wormatlas.org/neurons/Individual%20Neurons/AVBframeset.html), which are also in this arm.

The motor neurons in this arm consist of some VA neurons, and majority of AS neurons, which are responsible for locomotion.

The sensory neurons include PVDL, PVDR which responds to hard touch and cold temperatures. It being in this arms explains for a need to have faster signal transfer from PVD to motor neurons, since generally the worm needs to respond quickly to such strong stimuli.

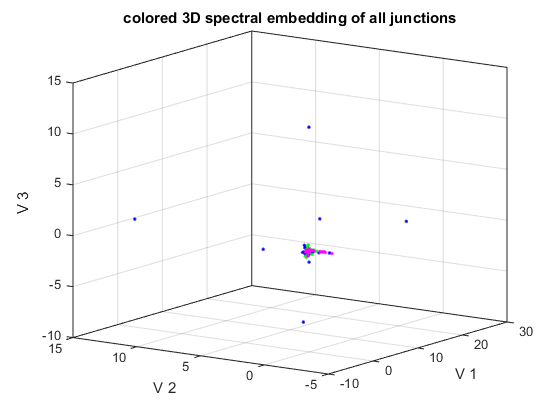
Hubs in the network:

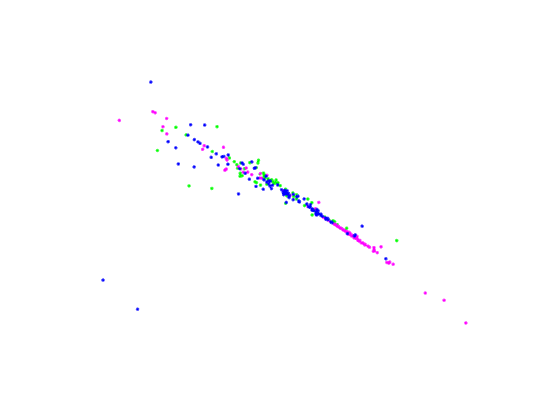
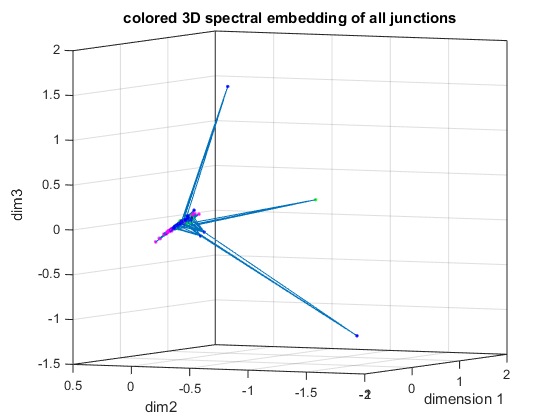
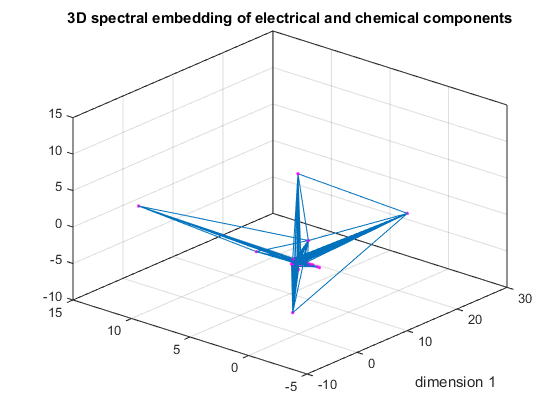
AVAL/AVAR: central cluster

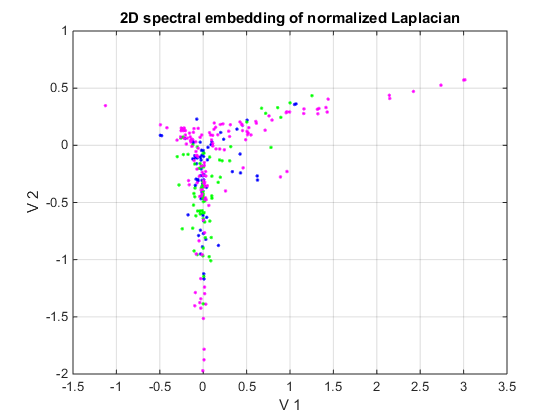
PVCL/PVCR: right arm

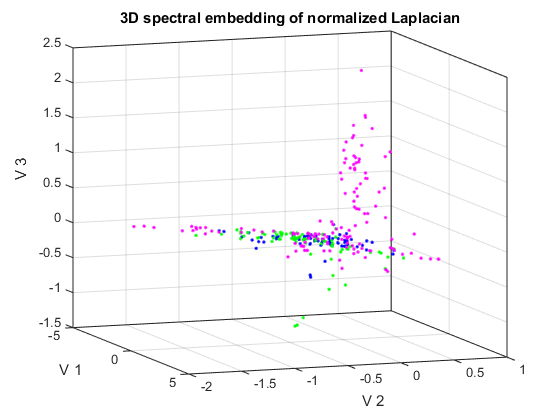
RIAL/RIAR: left arm

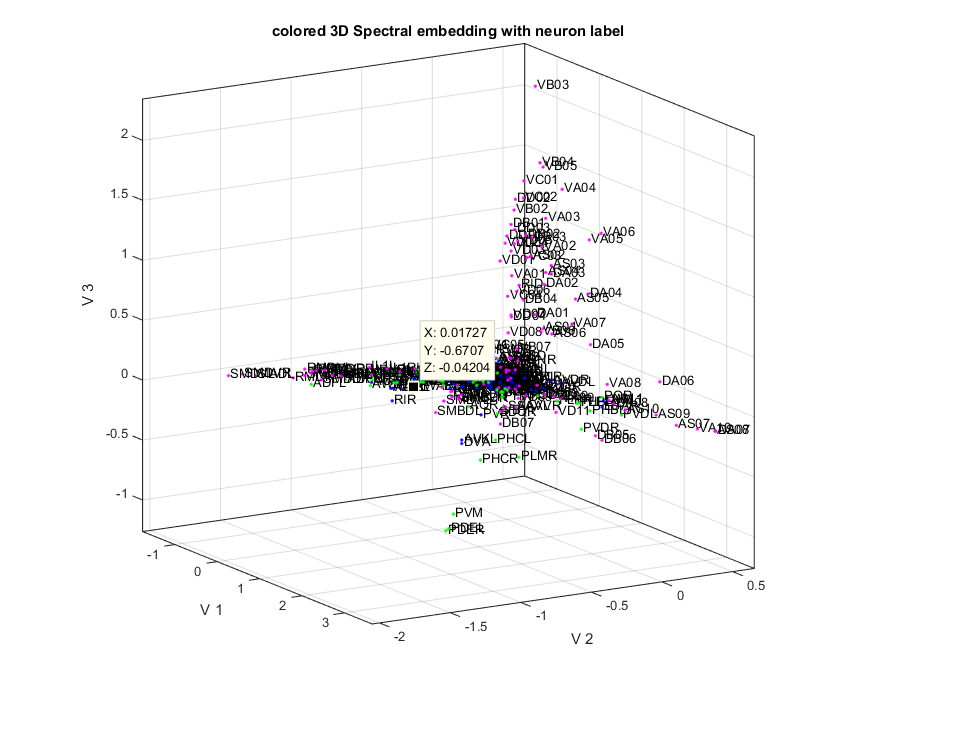
DVA: bottom arm

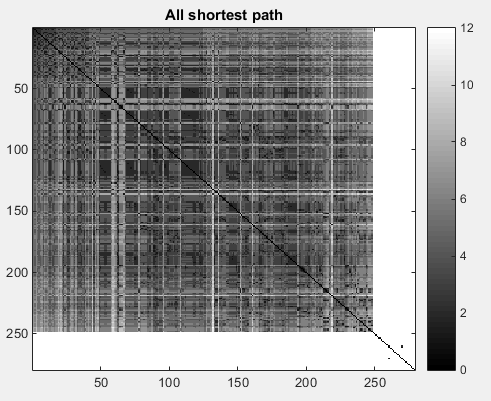












# Methodology

## Advantages

This method is efficienct since only a constant number of eigenvalue needs to be computed for the embedding. Typically k-means is the next step in clustering after the embedding. ans can be used as a subsequent clustering algorithm after the embedding.

